

实验报告

10225501438 侯志远

一.项目说明

众所周知，股票一直以来都是大家愿谈其详的一个话题，懂行的凑近去玩个开心，不懂得在外面看个开心。但股票终究是有风险的，在股市当中，既有像巴菲特那样利用人脉和自身丰富的专业知识和经验混的风生水起的，也有贪图一时蝇头小利而赔尽所有的可怜人。这项报告这是深入去看看股市当中一些基本数据的波动，走向，利用数据处理与可视化让我们了解一些股市表面的一些东西。

二.数据处理和导入

选取 A 股作为背景，对一些热门行业以 2018 年到 2023 6 月份（一个适中的数据）进行爬取和处理

数据来源于下图所示

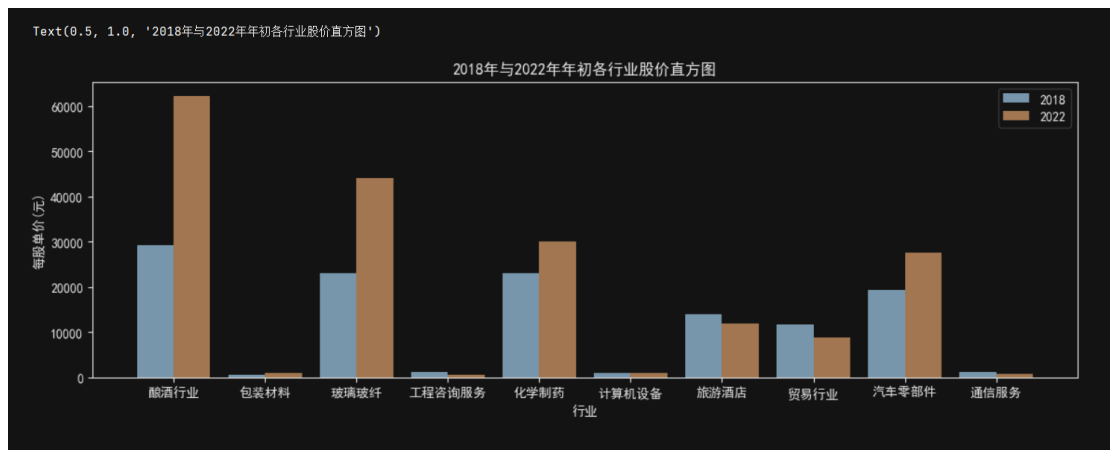


数据的处理如下

```
count = 1
cur_dir = os.path.dirname(__file__)
for name in arr:
    df = pd.read_excel(f'C:\Users\houzy\Desktop\finalhomework\数据\行业数据(周)\{hy}\{name}', sheet_name=0, index_col=0,
                      dtype={'股票代码': 'str', '日期': 'str'})
    for j in range(2018, 2023):
        for i in range(2, df['日期'].size + 2):
            if int(df['日期'].str.split('-')[i][0]) == j:
                kp = df['开盘'][i] * 1000 + kp
                sp = df['收盘'][i] * 1000 + sp
                zg = df['最高'][i] * 1000 + zg
                zd = df['最低'][i] * 1000 + zd
                zdf = df['涨跌幅(%)'][i]*1000+zdf
                zde = df['涨跌幅'][i] * 1000 + zde
                cjl = df['成交量'][i] * 1000 + cjl
                cje = df['成交额'][i] * 1000 + cje
                zf = df['振幅(%)'][i] * 1000 + zf
                hsl = df['换手率(%)'][i] * 1000 + hsl
```

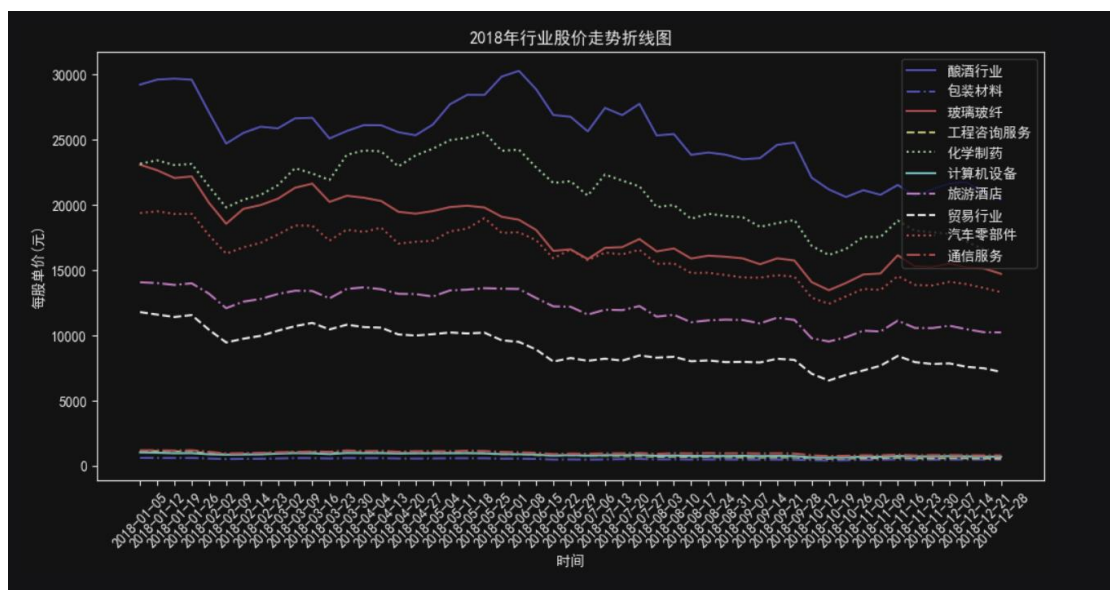
股票代码	股票简称	日期	开盘	收盘	最高	最低	涨跌幅(%)	涨跌幅	成交量	成交额	振幅(%)	换手率(%)	行业	地区	概念1	概念2	概念3	概念4
000012	南玻A	2018-01-06.14	6.36	6.44	6.13	2.25	0.14	382426	32735907	4.98	2.53	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.33	6.57	6.80	6.30	3.30	0.21	678591	60245448	7.86	4.50	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.49	6.46	6.53	6.36	-1.67	-0.11	400037	34932401	2.59	2.65	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.41	6.44	6.58	6.41	-0.31	-0.02	309424	27219840	2.63	2.05	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.49	6.58	6.66	6.43	2.17	0.14	404406	35964132	3.57	2.68	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.64	6.60	6.73	6.56	0.30	0.02	418783	37588139	2.58	2.77	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.52	6.46	6.60	6.39	-2.12	-0.14	281629	24715824	3.18	1.87	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.50	6.57	6.71	6.48	1.70	0.11	363202	32393734	3.56	2.41	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.56	6.42	6.63	6.40	-2.28	-0.15	246510	21658186	3.50	1.63	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.42	6.44	6.58	6.40	0.31	0.02	264719	23281507	2.80	1.75	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.38	6.53	6.56	6.36	1.40	0.09	236309	20670553	3.11	1.57	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.51	6.46	6.53	6.38	-1.07	-0.07	273388	23827826	2.30	1.81	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.49	6.68	6.71	6.45	3.41	0.22	521935	46750972	4.02	3.46	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.66	6.71	6.94	6.56	0.45	0.03	557742	50908680	5.69	3.70	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.63	6.80	6.85	6.60	1.34	0.09	346497	31535628	3.73	2.30	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.80	6.75	6.82	6.69	-0.74	-0.05	218889	19939853	1.91	1.45	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.71	6.74	6.75	6.61	-0.15	-0.01	288370	26029886	2.07	1.91	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.74	6.64	6.74	6.64	-1.48	-0.10	205457	18525214	1.48	1.36	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.63	6.81	6.84	6.61	2.56	0.17	387873	35360860	3.46	2.57	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.79	6.66	6.83	6.60	-2.20	-0.15	308470	27986256	3.38	2.04	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.68	6.53	6.70	6.49	-1.95	-0.13	213692	18970786	3.15	1.42	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-01-06.48	6.44	6.57	6.37	-1.38	-0.09	196877	17291639	3.06	1.30	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-06.42	6.15	6.49	6.12	-4.50	-0.29	275804	23484772	5.75	1.83	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-06.09	6.23	6.26	6.00	1.30	0.08	177508	14813282	4.23	1.18	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-06.15	6.24	6.27	6.08	0.16	0.01	120477	10138767	3.05	0.80	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-06.07	5.73	6.16	5.71	-8.17	-0.51	287491	23394361	7.21	1.90	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-05.89	5.73	5.96	5.68	0.00	0.00	181408	14437434	4.89	1.20	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-05.76	5.82	5.88	5.76	1.57	0.09	119249	94575394	2.09	0.79	玻璃玻纤	广东	广东板块	军工	新能源	AB股	
000012	南玻A	2018-02-05.70	5.50	5.73	5.42	-5.50	-0.32	218334	16659298	5.33	1.45	玻璃玻纤	广东	广东板块	军工	新能源	AB股	

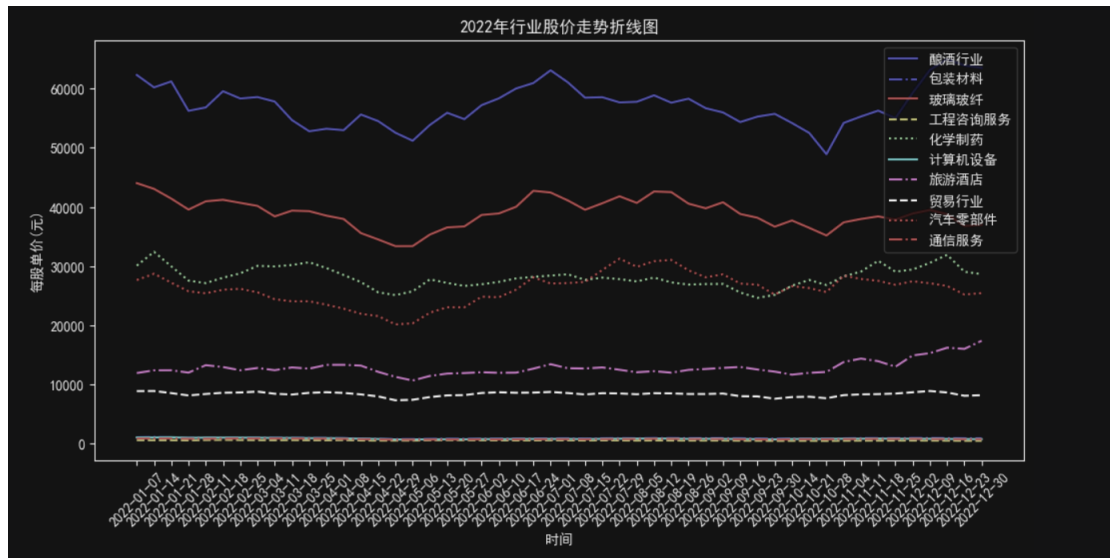
上述处理完之后，我们便可以开始导入数据，绘制成比较形象的图来直观的看这些变化



通过上面这副图我们能够 直观的看到 2018 和 2022 年各行业股价的自身对比情况，大部分行业相比 2018 年，2022 年股价有明显上升，但少数如同贸易，旅游等受到限制，这不由得可以联系到疫情对这种行业产生的影响，在大方向来看，若是考虑到疫情来说，居家隔离的股民选择将资金投入到了制药，酿酒，这种可以想到的行业，应该会小小的收获一笔。

那我们当然不能只看大年份的变化, 下面便是疫情前两年(2018)和疫情后两年 (2022) 的单年行业股价走势折线图



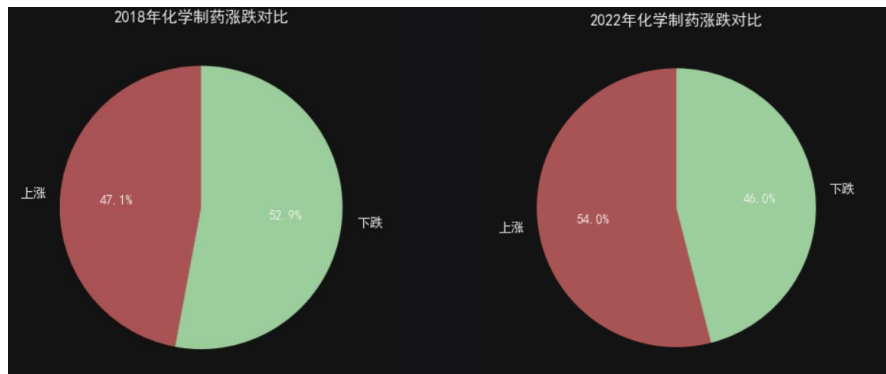


由图表数据可以看出 2018 年中国 A 股大部分行业均出现下行，其中中国本土传统行业均出现大幅度下行。

2018 年是一年内外部环境变化较大，加之贸易战等政治经济风险因素的影响，A 股市场整体表现不佳。

由图标可以看出 2022 年中国 A 股大部分行业处于区间震动,在年末,由于对新冠疫情的管控放松,行业股价出现了两极分化,有些行业迅速拉升,如酿酒行业,旅游酒店。而有些行业极具下跌,如化学制药。我们看到了,即使两者有着疫情的差别,股价的变化还是很难预测的波动性,在没有绝对的了解情况下,贸然买入,还是会产生不可预料的结果的。

下面这是化学制药 2018 年和 2022 年两年各自的涨停对比



从上图中可以看出，2018 年化学制药行业的涨跌幅比较接近，下跌幅度略高于上涨幅度，整个行业整体呈现出下跌的趋势。而 2022 年，化学制药行业整体上涨，且上涨幅度高于下跌幅度，表明化学制药行业正在逐渐获得市场的认可和支持。

三.总结与反思

爬取数据方面：在爬取的时候直接统一单位，可以大大减少后面数据处理出现的问题

建立模型方面：在建模前，首先要了解数据的类型、结构和内在关系。这包括识别数据的尺度（比如名义、序数、区间、比例）、分布和潜在的异常值,最好选用合适的图例来建立模型来达到直观美感的效果。

这次实验带来的直观感受：股票市场具有一定的周期性，且与宏观经济因素密切相关建议继续对历史数据进行深入分析和挖掘，结合宏观经济因素，预测未来股票市场的走势，从而制定更加合理的投资计划。